

2.5D U-Net with Dense/Residual Layers and Global Context Blocks for Kidney Tumor Segmentation

Lin Qi¹, Shuangwei Liu², Ying Wei³

^{1,2,3} Northeastern University, Shenyang 110819, China
q731787474@gmail.com

Abstract. To address the segmentation of kidney and kidney tumor segmentation on CT scans, we propose a new 2.5D convolutional neural network that can utilize global context and multi-stage features.

1 Method

We use a 2.5D UNet [1] as the baseline, specifically, we fill the channel axis with a volume frame around the input slice to combine semi-3D information [2]. We use two different 2.5D U-Nets to segment kidney and kidney tumors, and combined these two results to form the final prediction. The encoder takes the 2.5D input patches extracted from the set of original CT images. For kidney segmentation, the patch size is heuristically defined as $3 \times 256 \times 256$ voxels on the z, y and x coordinates by considering the spacing information. For kidney tumor segmentation, the patch size is $3 \times 128 \times 128$.

Using the UNet as a baseline, we adjust the depth of the network and the number of filters, and extended the basic convolution module into a dense connected module [3], using summation instead of concatenation. We also added a global context (GC) block after each convolution module [4].

The method of spatial attention is introduced into the GC module, which can directly calculate the interaction between any two locations and capture the remote dependency. At the same time, the channel attention is also used in the GC module, and the characteristics of different channels are reweighted by using the features with spatial context information. In this way, important features are promoted and unimportant features are suppressed.

Unlike the original U-Net, we use a convolution layer with stride of 2 instead of maxpooling for downsampling. Inspired by [5], we use depth supervision in the decoder, integrate the segmentation layers at different levels of the network, and combine them through the summation of elements to form the final network output.

In addition, when training the kidney segmentation network, the cross entropy loss is used, and the dice loss combined with the cross entropy loss is used for the kidney tumor segmentation network.

All convolutional layers have 3×3 kernels and leaky ReLU activation and a batch normalization layer [6] is used after each convolutional layer.

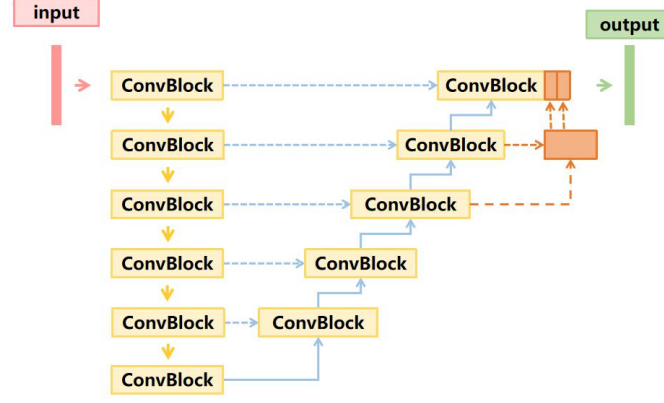


Fig. 1. Network architecture. The network is improved by U-Net. The encoder aggregates the high level information, and then the decoder uses this information to locate precisely. Inspired by [5], we inject gradient signals into the depths of the network through depth supervision.

2 Implementation Details

To address the intensity variations between subjects, we normalize the intensities in each slice so that the mean of intensities equals to zero and the standard deviation equals to one for each modality.

The model is trained with Adam optimizer. The learning rate is initialized to $3e-5$, after 25 iterations, the learning rate is reduced to $1/10$ of the previous one.

References

1. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Proceedings of Medical Image Computing and Computer Assisted Intervention. (2015).
2. Xu Y, Géraud T, Bloch I. From neonatal to adult brain MR image segmentation in a few seconds using 3D-like fully convolutional network and transfer learning[C]//2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017: 4417-4421.
3. Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.Author, F.: Contribution title. In: 9th International Proceedings on Proceedings, pp. 1–2. Publisher, Location (2010).
4. Cao Y, Xu J, Lin S, et al. GCNet: Non-local Networks Meet Squeeze-Excitation Networks and Beyond[J]. arXiv preprint arXiv:1904.11492, 2019.
5. Kayalibay B, Jensen G, van der Smagt P. CNN-based segmentation of medical imaging data[J]. arXiv preprint arXiv:1701.03056, 2017.

6. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[J]. arXiv preprint arXiv:1502.03167, 2015.